利用视觉情境范式揭示口语加工的时间进程

魏一璞　北京大学对外汉语教育学院

摘要 视觉情境范式是一种通过追踪、测量人眼在视觉物体上的注视轨迹来研究实时口语加工的眼动实验范式。该范式运用于语言理解类研究的理论基础是眼动连接假设（如：协同互动理论、基于目标的连接假设理论等），这些连接假设在眼动轨迹与口语加工进程之间建立起了有意义的关联。使用视觉情境范式所获取的数据能够为口语加工提供精确的时间信息，常用的数据分析方法包括：时间兴趣区内注视比例均值分析、分叉点分析、生长曲线分析等。该范式为研究词汇语音识别、句法解歧、语义理解、语篇语用信息加工等问题提供了关键性证据。

关键词 视觉情境范式，眼动追踪，口语加工

语言加工的时间进程问题一直是心理语言学领域的核心议题之一。探讨这一问题有三层重要意义：首先，不同层面的语言信息(语音、语义、句法、语篇、语用等)以及不同来源的信息(语言输入、视觉环境、世界知识等)在何时被认知系统加工处理对于语言理解模型的建构至关重要。例如，McRae 等人(1998)的基于约束的语言加工模型(constraint-based model)，就是根据歧义句理解的时间进程证据提出的。其次，研究影响语言理解的因素(如：词频、语言水平、认知能力等)如何起作用也需要语言加工的时间进程信息，如：Magnuson 等人(2003)通过考察听者理解语音输入时锁定目标指代对象的时间，提出了词频会影响词汇识别的论断。此外，语言要素加工的时间线也可以作为反映语言理解能力的重要指标，在儿童语言习得、二语加工以及老年人语言能力评估等方面发挥作用(Saryazdi & Chambers, 2021)。作为探究口语加工时间进程的重要工具，视觉情境范式(visual world paradigm)可以提供精确的时间信息，揭示各个层面口语加工的问题。

视觉情境范式是一种通过追踪、测量人眼在视觉环境中的注视轨迹研究实时口语理解加工的实验范式(Allopenna et al., 1998; Salverda & Tanenhaus, 2018)。随着 60 年代末眼动追踪仪器与电脑接口的实现，即时记录眼动轨迹以及自动处理眼动数据成为可能。70 年代中期，利用眼动技术进行的书面阅读研究已取得大量进展(综述见：Rayner, 1978)。与此同时，Cooper

收稿日期：2023-02-06
通讯作者 Email：weiyipu@pku.edu.cn
基金支持：教育部人文社会科学研究青年基金项目（21YJC740062）

(1974)第一次尝试使用眼动追踪技术对口语理解进行测量。这项早期研究首次将听者对视觉物体的注视与语言加工建立了联系。1995 年 Tanenhaus 等人在《科学》上发文，阐释了如何利用眼动追踪技术揭示歧义句的加工过程，视觉情境范式(由 Allopenna et al., 1998 定名)才开始大量被应用于口语加工研究，成为心理语言学、认知心理学领域最重要的研究手段之一(邱丽景等, 2009; 林桐, 王娟, 2018)。

本文主要阐释了如何利用眼动视觉情境范式探究口语加工的时间进程。为了阐明这一问题，本文将首先介绍眼动实验范式中的连接假设，将视觉场景中的眼动轨迹与语言的理解过程建立起联系，并且充分说明视觉情境范式在任务及数据上有哪些时间性的特点，以及如何利用这些特点进行数据分析；进而以口语加工的时间进程为主线，综述近 20 多年来使用该范式的研究在语音、语义、句法、语篇与语用加工等方面的实证发现，进一步说明这一高时间敏感性范式在口语加工时间进程研究中的贡献。

1. 眼动轨迹与语言加工进程的连接假设

视觉情境范式眼动研究方法的理论基础是连接假设(linking hypotheses)，这类假设将眼动轨迹与口语理解的认知过程建立起了联系(Allopenna et al., 1998; Tanenhaus et al., 2000)。具体来说，当听者处理口语信息时，会将语言所描述的情景形成动态的心理表征(mental representation)；而理解者对心理表征中特定实体的关注会随着语言信息的输入而变化——相应地，他们在视觉空间中的注视点也会随之移动(Altmann & Kamide, 2007)。这种注视的聚集和移动，伴随着瞳孔位置的改变。通过眼动追踪手段，瞳孔位置移动的轨迹可以被有效测量，进而揭示口语加工的时间进程。过去二十年间，学界提出了诸多反映眼动与口语加工之间关系的具体连接假设理论，用于阐释视觉注意如何被分配到指代物体之上(见综述 Magnuson, 2019)。本文总结了其中比较有影响力的三种连接假设理论，进一步阐明将视觉情境范式应用于口语加工研究的理论基础。这些连接假设虽未直接就具体语言元素加工的时间进程进行界定，但其假设中包含了口语加工的若干阶段，是探讨加工时间进程的前提基础。

Knoeferle 和 Crocker (2006, 2007)提出的协同互动理论(coordinated interplay account)将基于视觉情境的口语理解分为三个主要阶段：(1)在原有的语句结构中整合新输入的词，形成新的语句理解，并基于这一新信息和原有的语言信息、相关世界知识，共同形成对后面语句的预测；(2)在包含之前视觉场景的工作记忆中，搜寻词语所指代的物体或者是基于第一阶段信息可以预测到的物体；(3)将语言输入(名词、动词等)与视觉场景中的物体、动作对应，

基于视觉场景信息修正之前形成的语句理解，并形成新的预测(Knoeferle & Crocker, 2006, 2007; Pyykkönen-Klauck & Crocker, 2016)。值得注意的是，这三个进程虽然在协同互动理论中依次呈现，但该理论并不排斥三个进程在加工时间上有交叠或者同时发生的可能性。协同互动理论凸显了视觉场景信息对于口语理解的重要性；而且尽管当视觉场景消失后，这些情景在工作记忆中会逐渐消退，但关于情景的记忆仍然对后续句子加工具有显著的影响(Knoeferle & Crocker, 2007)。

Altmann 和 Mirković (2009)提出了另一种连接假设理论，这一理论同样也认同语句加工受到语言信息(如：实时语言输入、语境信息)和非语言信息(如：视觉场景、世界知识)的共同影响。但不同于 Knoeferle 和 Crocker (2006, 2007)的协同互动理论，Altmann 和 Mirković (2009)认为处理视觉场景信息与理解语言输入的过程在心理表征和处理时间上都是无法分割的——因为语言信息和非语言信息都存储在同一套系统中，共同构成了对情景的动态表征。当听者接收到某一信息时，关于客体的表征(包括与此客体相关的体验、知识等)会被激活。而随着听者不断接收不同来源的信息(语言输入、视觉场景、世界知识等)，关于客体的表征就会不断变化。当不同来源的信息出现重合时，客体表征的激活就会加强。这一表征系统的不同状态体现在心智表征(mental representation)层面就是注意力的分配，而注意力的分配影响了眼动轨迹。换言之，伴随语句输入，受试者对视觉物体的注视在时间上的变化轨迹，是由包含语言信息、语境信息、视觉场景、世界知识等的一套共同表征系统所影响并驱动的。在该理论假设框架下，不同来源的信息对口语加工会产生即时影响，也会迅速反映在眼动轨迹上。

以上两种连接假设均基于语言理解视角，将口语加工过程中的眼动注视变化看做是语言输入信息与视觉信息共同作用的结果。这两个假说都将语言加工看做是一项独立的任务，与实验过程中的行为任务目标无关。然而，此类基于语言理解视角的连接假设未涉及完成任务所需要的动作本身对语言指代加工的影响(Chambers et al., 2004)，同时也未考虑到在视觉搜寻中眼动本身就和行为任务的目标紧密相关——即受试者会更多地注视与自己行为目标相关的物体。为了更好地解释语言加工与眼动的关系，Salverda 等人(2011)提出了基于目标的连接假设理论(goal-based linking hypothesis)，将“任务目标”这一新维度纳入眼动连接假设。不同于基于语言理解视角的连接假设，基于目标的连接假设理论认为不仅语境、语言输入等可以对语言加工形成约束(constraint)，任务目标本身也可以作为约束——与执行任务目标直接相关的视觉物体，会吸引更多眼动注视；而与目标执行无关的物体则不会。该连接假设理论认为，视觉情境下的口语加工过程首先包含了一项基础任务，就是把语言输入信息与视觉

场景中可供选择的物体对应，而眼动注视服务于这一任务目标，用于锁定可能的指代物体；不符合可供性(affordance)的物体则很少被注视。例如，在听到 *put the cube into the can* 这一指令时，只有尺寸大小能放得下立方体(cube)的罐子(can)才会成为被注视的目标容器(Chambers et al., 2004)。Salverda 等人(2011)认为，额外的任务如点击物体、移动物体等，共同构成了口语加工任务中的任务目标结构，并且影响了眼动注视。例如，当受试者带着判定句子正误任务听句子时，会比无判定任务情况下听同样的句子展现出更早、更显著的预测性注视(Altmann & Kamide, 1999)，在时间进程上更快地锁定指代目标。基于目标的连接假设为细化、层级化语言加工过程中的任务目标结构提出了新的要求。

利用眼动视觉情境范式进行的口语加工研究以连接假设为基本前提，根据利用视觉信息的情况，可以分为两个主要研究方向。第一类研究将视觉场景作为呈现物体的布景，心理表征中对特定指代对象的注意被投射在视觉场景中，听者据此形成对指代物体的注视；而其注视布景上的物体所形成的眼动轨迹，揭示了不同的语言成分如何被实时加工 (例如：Cooper, 1974; Cozijn et al., 2011; Kaiser, 2016)。第二类研究则将视觉信息也作为一种语境约束，主要探索视觉环境中的信息(如：候选物体个数、物体大小对比、所描绘的事件动作等)本身对语言加工产生的影响(例如：Chambers et al., 2002; Knoeferle et al., 2005; Tanenhaus et al., 1995)。这两类研究采用的任务类似，但是在连接假设的理论层面，第一类研究强调了理解视觉场景信息与理解口语输入信息这两个过程的共时性和不可分割性；第二类研究则将视觉场景信息加工作为一个相对独立的过程，强调视觉场景本身在口语加工过程中的作用。而眼动加工领域最新的趋势是开始关注任务目标对语言加工的潜在作用。尽管纳入了目标维度的连接假设已经完成了初步的理论建构，但目前针对不同任务目标下加工效应对比的研究仍然是空白。

2. 视觉情境范式的特点

2.1 范式与任务

典型的视觉情境范式实验通常包含以口语形式呈现的语言指令和以视觉刺激形式出现的物体(在真实世界中或者电脑屏幕上)。受试者在理解口语指令的同时，在视觉物体上注视点的位置被眼动仪实时记录并用于后续分析(见图 1)。视觉刺激图片一般会先于语言指令出现，并有一定的预视时间；语言指令以相对固定的播放速度呈现。前人研究中发现，图片复杂度、预视时长、语言指令播放速度、任务指令类型(是否明确告知受试者需要预测目标物)等因素都会对眼动结果产生一定的影响(Huettig & Guerra, 2019; Ferreira et al., 2013)。

视觉情境范式主要包括两种不同的实验任务：一是主动任务(基于动作的实验任务)，即

要求受试者对语言指令做出行为上的反应(如：获取、挪动、点击物体；见 Hanna & Tanenhaus, 2004; Tanenhaus et al., 1995)；二是被动任务(听-看任务)，即受试者仅需要听语言指令、看图片或者情景，不需要在行为上做出反应(Altmann & Kamide, 1999; Knoeferle et al., 2005)。关于两种任务的区别，Salverda 等人(2011)指出在主动任务型视觉情境范式实验中，获取、挪动、点击物体之前受试者会将大量的注视投向目标物体；而被动任务型实验不存在这样的注视模式——此因素可能会导致两种实验任务下眼动模式的差异。Pyykkönen-Klauck 和 Crocker(2016)综述对比了采用两种任务类型的眼动实验结果，认为主动任务中一些语言效应(如：词频效应)在眼动指标上表现得更为敏感，受试者能更快地锁定目标物体，显示出更迅速实时的语言理解过程。而听句子看图的被动任务型视觉情境范式实验，因不需要受试者完成额外任务，相对而言具有更好的生态效度(Huettig et al., 2011a)；而且可以被用于检验哪些口语加工效应是在语言与视觉交互中普遍存在的，哪些仅在特殊的实验任务下才存在(Huettig et al., 2011b)。



图 1 视觉情境范式实验呈现示例

视觉情境范式有两个主要的变体——拼词呈现范式(printed-word paradigm，Huettig & McQueen, 2007)与空屏呈现范式(blank screen paradigm，Altmann, 2004)。拼词呈现范式中，视觉刺激图片被替换为出现在屏幕上的词语。受试者会听到与该词相关的语音输入，同时其在每个字母上的眼动注视轨迹被记录下来用于分析。拼词呈现范式可以用于检验语音的识别过程、研究正字法信息如何被实时加工等问题。空屏呈现范式主要用于揭示短期记忆在实时语言加工中的作用。在视觉刺激图片呈现几秒后，呈现空白屏幕(一般 1 秒)，然后播放语音

指令。采用该范式的实验可以证明，即使在视觉刺激图片中的物体消失之后，受试者听到语言指令仍然会看向相关物体原来所在的位置(Knoeferle & Crocker, 2007)。空屏呈现范式为心智表征提供了依据：心智表征形成后，可以不依赖视觉刺激，而暂时存储在短期记忆中，参与后续的语言加工。

2.2 数据与变量

视觉情境范式实验数据分析中的常用因变量为注视和眼跳。其中最常用的注视指标是注视比例(fixation proportion)，即在指定时间窗口内落入某一兴趣区的注视点在所有试次中的比例。眼跳(saccade)数据常用的指标包括眼跳比例(即所有试次中看向目标兴趣区的眼跳比例)和眼跳反应时(即当目标词刺激出现后，看向目标兴趣区所需要的眼跳时长)。数据中的自变量可以是实验设计的组内变量(如：实验条件与控制条件、歧义句与非歧义句等)，也可以是组间变量(如：不同语言背景组、年龄组等)。

视觉情境范式的优势在于所产出的数据具有高度的时间精确性，现有的科研用眼动仪可以达到 1000Hz 的取样率，即每一毫秒捕捉一次眼动位置，可以提供准确的时间进程信息。以兴趣区注视比例这一数据指标为例，研究者不仅可以跨组对比在某一时间窗口内不同条件组下注视比例的均值，以确定口语加工中的某一效应；更重要的是可以探究效应出现的时间(即注视比例在不同条件下开始产生显著区别的时间)以及效应随着时间发展而变化的曲线模式。

2.3 利用时间维度信息进行数据分析

时间上的精确性是视觉情境范式数据的最重要特点，如何利用好时间维度信息是该范式数据分析的关键。根据利用时间信息的方式，可将现有的数据分析方法归为三类：(1)指定时间兴趣区内注视比例均值对比；(2)效应出现、持续的时间进程分析；(3)效应随时间变化的曲线模式分析。为了更好地阐释三类方法的应用场景与分析逻辑，本文选用了 Allopenna 等人(1998)研究中的实验物体示意图(图 2)和注视比例数据图(图 3)作为示例(该研究的详细讨论见第 3.1 节)。

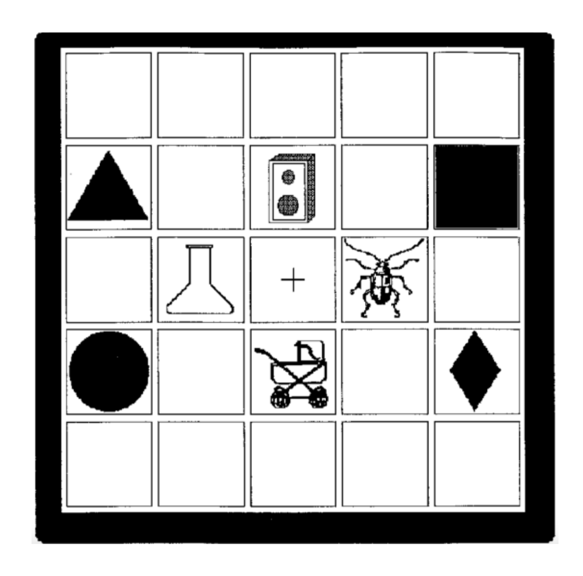


图 2 视觉情境范式实验视觉刺激示意图。语音指令为：*beaker*“烧杯”。四个用于测量的物体分别为：左-目标指代物体(referent) *beaker*“烧杯”、 右-语音同群竞争项(cohort) *beetle*“甲虫”、上-韵律竞争项(rhyme) *speaker*“扬声器”、下-无关项(unrelated) carriage“婴儿车”。资料来源：Allopenna 等人(1998)，已获使用许可。

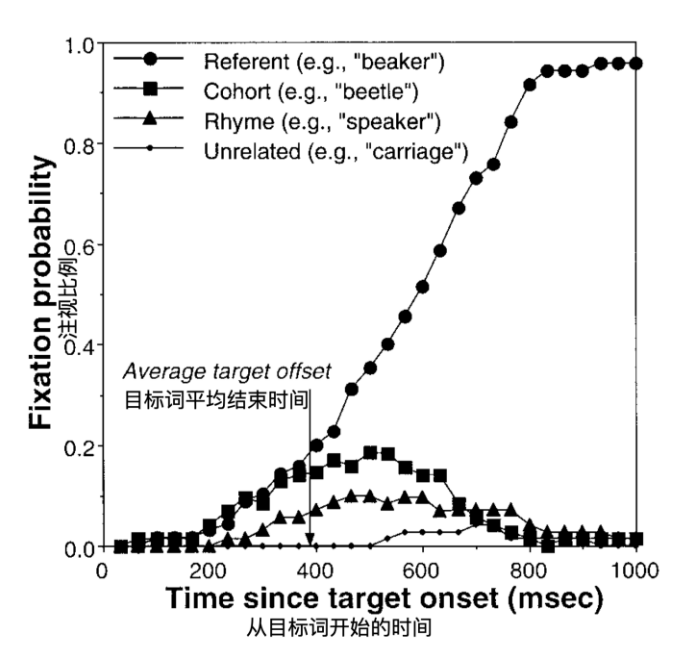


图 3 使用视觉情境范式下词汇识别任务所得数据示意图。横轴：从目标词开始呈现后的 1000 毫秒时间轴；纵轴：注视比例。四条曲线分别代表看向目标指代物体(referent) *beaker*“烧杯”、语音同群竞争项(cohort) *beetle*“甲虫”、韵律竞争项(rhyme) *speaker*“扬声器”、无关项(unrelated) carriage“婴儿车”的注视比例。资料来源：Allopenna 等人(1998)，已获使用许可。

第一类分析方法是分析视觉情境范式数据最常用、最直观的方法——将指定时间兴趣区

内注视比例均值进行对比，如：对比从目标词 *beaker*“烧杯”开始呈现到目标词结束的约 375 毫秒内听者对图 2 中几个物体的注视比例。这一分析方法将注视比例、时长或者眼跳指标作为因变量，组内和组间变量作为自变量，采用 t-test、ANOVA、混合效应模型(linear mixed-effects model)等统计手段对比不同物体之间或者不同条件组之间注视比例的差别。相比于 t-test 和 ANOVA，混合效应模型是目前应用最广的分析方法，它可以将受试者之间以及试次之间的差异作为随机变量纳入模型，实现对效应更准确的模拟与测试(应用示例：Gardner et al., 2021; Grüter et al., 2020)。需要注意的是，此类统计方法通常要求数据符合正态分布，而注视比例的阈值范围在 0 到 1 之间，一般需要事先进行对数(log)转换或者逻辑(logit)转换(Ito & Knoeferle, 2022)。分析指定时间兴趣区内注视比例均值是最简便的视觉情境范式数据分析方法，适用于大部分实验设计。其主要劣势在于人为设定的时间窗口降低了数据的时间精度，无法很好地捕捉注视比例随着时间变化的趋势；补偿方法可以是将不同时间兴趣区作为自变量加入分析模型，检验时间兴趣区这一变量本身是否显著影响注视比例。

第二类方法是对效应出现、持续的时间进程进行分析。此类方法充分利用了视觉情境范式精确的时间进程信息，可用于探究某一口语加工效应出现的确切时间。其中，分叉点分析(divergent point analysis)将潜在效应出现的时间段再细分为若干小的时间窗口(如 20 毫秒)，在每一个小的时间窗口内对比检验两个条件组的注视比例是否存在显著区别，从而找出两组注视比例曲线最早开始出现显著分叉的时间点。例如：图 3 中目标指代物体 beaker“烧杯”的注视比例曲线与语音同群竞争项 *beetle*“甲虫”的注视比例曲线分叉点大约在 400 毫秒左右，晚于目标指代物体与韵律竞争项 *speaker*“扬声器”的注视比例分叉点，而通过分叉点分析可以统计计算出不同曲线之间开始显著分叉的具体时间点。

简单的分叉点分析只能界定出效应开始的时间点(两个条件下变化曲线的分叉点)，并不能检验分叉点在时间上的变化区间，也不能跨条件组比较两个分叉点是否存在统计学意义上的显著不同。而基于自助抽样检验(bootstrapping)的进阶分叉点分析法，则可以为每一个分叉时间点提供置信区间，从而实现跨条件组对比(Stone et al., 2021; 应用示例：Corps et al., 2021)。进阶的分叉点分析法可为对比研究不同群体语言实时加工的时间进程提供有效的分析工具，例如，一语者与二语者在语言加工的某一效应上(如：预测加工)可能并不存在效应量上的差别，但是两类被试者在效应开始的时间上可能存在差异(Kaan & Grüter, 2021)，采用这种分析方法就可以有效检验二语者预测加工开始的时间是否会显著地滞后于一语者。除了分叉点分析法，基于频率簇的置换检验法(cluster-based permutation analysis，Barr et al., 2014)和自助抽样检验时间序列差别法(bootstrapped differences of timeseries，Seedorff et al.,

2018)，均可以用于界定两个条件组数据出现显著差别的时间(详见眼动数据分析方法综述：Ito & Knoeferle, 2022)。但此类分析方法均无法对不同条件下效应随时间变化的趋势进行分析，要回答此类问题需要借助第三类方法分析变化曲线。

第三类方法主要针对视觉情境范式中效应随时间变化的曲线模式进行分析。其中，生长曲线分析法(growth-curve analysis)将不同条件组下关键兴趣区的注视比例随着时间变化的曲线进行模拟、分析，检验不同条件组下注视比例曲线变化的模式是否有所不同，进而检证效应是否随着时间发展有所变化(Mirman, 2014; Mirman et al., 2008)。不同于第一类分析法，生长曲线分析法不仅包括了以时间作为变量的线性模型，还可以在模型中加入时间变量的二次方、三次方，以模拟注视比例随着时间出现曲线变化的模式[1]，如在图 3 中对语音同群竞争项 *beetle*“甲虫”的注视比例出现了呈抛物线状先升后降的趋势，且斜率不同于韵律竞争项 *speaker*“扬声器”，这一模式就可以采用包含二次方时间变量的生长曲线模型进行分析。在口语加工过程中，注视随着时间的变化趋势常常并非线性上升或者下降，对变化曲线的模拟和对比能够更精确地分析语言理解的时间发展进程(应用示例：Henry et al., 2022; Koring et al., 2012; Wei et al., 2019)。需要注意的是，生长曲线分析法存在数据自动相关性问题(autocorrelation)，即相邻的两个时间窗口在注视位置上存在高度相关性，增加了出现统计学一型错误(假阳性)的几率(Huang & Snedeker, 2020)，因此常需要与第一类和第二类的分析方法相结合，共同验证效应。广义加性混合模型(generalized additive mixed model)分析也可以用于对非线性的数据曲线进行模拟，通过薄板样条插值(thin plate regression splines)更灵活地模拟变化曲线，并且减少统计学上的自动相关性，一定程度上弥补了生长曲线分析法的劣势(Porretta et al., 2018)。

3. 视觉情境范式与口语加工的时间进程

学界早年关于语言加工时间进程的争论主要集中在加工即时性问题上。早期实验主要采用词汇再认、线索回忆、自定步速阅读等任务，得到的证据倾向于支持延迟整合加工(如：Garnham et al., 1996; Stewart et al., 2000)，即语言使用者加工语言会延迟到句子末尾再进行整合(delayed-integration interpretation; Millis & Just, 1994)。然而，随着眼动、脑电事件相关电位(ERP)等测量方法的推广，精确测量阅读时间、脑电信号反应成为可能，越来越多的证据支持语言加工的即时性，即语言使用者会随着语言的输入即刻处理遇到的信息(incremental

[1]在包含时间变量的生长曲线基本模型中(如：$Y = \beta_0 + \beta_1 \times Time$)，$\beta_0$为截距，表示当时间为零时(即开始时)注视比例(Y)的数值；斜率 β_1表示随着时间的推移，注视比例的变化趋势；如将时间的二次方($Time^2$)、三次方($Time^3$)加入模型中，即可以允许注视比例随着时间推移呈抛物线变化——$Time^2$可以模拟有一次趋势方向变化(如先升后降，或先降后升)的曲线，而 $Time^3$可以模拟含两次方向变化的曲线。

interpretation; Traxler et al., 1997; Cozijn et al., 2011; Koornneef & Van Berkum, 2006)。对于视觉情境下的眼动测量，尽管从接收到听觉语言信号刺激到做出眼动反应需要大约 200 毫秒(Matin et al., 1993; Saslow, 1967)，使用视觉情境范式的大量口语实验中仍发现了在测试词开始呈现后、下一词未开始之前眼动注视投向目标物的效应，说明语言使用者对口语中信息的处理是即刻发生的(详见 3.1~3.5 小节)。

在即时性加工被广泛认可的基础之上，近年来语言加工时间进程的讨论主要聚焦于语言使用者何时利用语境信息来理解语言。语言使用者可能在测试词出现的同时，即时地结合测试词的语义与前文语境进行加工；也可能在测试词出现之前、加工语境信息的过程中，对测试词的语音、语义甚至所处的句法结构提前进行预测性加工(expectation-based account; Levy, 2008)。在对预测效应的检测上，视觉情境范式相对于阅读范式、ERP 测量等方法具有明显优势(Huettig & Guerra, 2019)。大部分采用后者的研究只能在测试词出现的位置捕捉到由测试词语义与语境信息一致性所产生的效应；而视觉情境范式可以在关键词出现之前，更早地检验到语境对受试者在视觉场景中注视方式的影响，为口语的预测性加工提供了关键性证据。下文将重点分析视觉情境范式在语音、语义、句法、语篇与语用等不同层面如何回答语言加工的时间进程问题。需要说明的是，不同层面的信息在口语加工中并非独立，而是会相互影响(见综述：Kuperberg & Jaeger, 2016)；而本文出于利于分类总结的考虑，将各个层面单列综述。

3.1 词汇识别与语音预测

视觉情境范式中，听者听到一个词就会在视觉范畴内寻找指代的物体。基于这一特点，视觉情境范式可以用来检验词汇的识别过程，并且探究听者如何利用已有信息预测语音形式。Allopenna 等人(1998)利用该范式检验了在口语词汇的语音识别过程中，语音输入与词汇表征的匹配过程是否是渐进发生的。如果这个匹配过程在时间上是渐进的，那么可以预测目标指代物体 *beaker*“烧杯”的语音同群竞争项 *beetle*“甲虫”，会比 *beaker* 的韵律竞争项 *speaker*“扬声器”有更强的干扰效应(见图 2)，因为语音上 *beetle* 与 *beaker* 在词语的开头位置有重叠，而 *speaker* 与 *beaker* 的重叠发生在后期。Allopenna 等人的视觉情境范式眼动实验结果验证了这一假设：注视目标物体“烧杯”的比例和注视“甲虫”的比例在语音加工的早期都出现了上升(见图 3)，而对“扬声器”这一物体的注视比例则是在词加工的较晚时间才出现上升，而且注视比例上升的幅度也相对比较小。视觉情境范式提供的眼动注视比例数据有效揭示了词汇识别中语音输入和词汇表征的匹配过程。

在语言使用者能否通过语境信息预测即将出现词语的语音信息这个问题上，已有的 ERP

研究结果存在很大分歧，并未能得到稳定可复制的语音预测效应(DeLong et al., 2005; Nieuwland et al., 2018)，而视觉情境范式为探讨语音预测问题提供了有力的证据。Ito 等人(2018)采用视觉情境范式的眼动实验，发现在高度可预测的语境下(例如：*The tourists expected rain when the sun went behind the...*)，听者不仅会预测性地注视目标物体(*cloud*“云”)，还会更多地注视目标物体的语音竞争项(与 *cloud* 共享开头音节的 *clown*“小丑”)，这一发现证实了语音形式预测的存在。更重要的是，在视觉情境范式下这一预测效应在目标词出现前的 500 毫秒就已经出现，充分证明语言加工中对语音形式的预测是主动的(proactive)，相比于一些其他范式仅在目标词位置发现整合效应的结果，视觉情境范式为语言预测提供了更为直接的证据。此外，视觉情境范式还为研究语音预测机制提供了实证依据：语音预测与语义预测一样，其背后机制都是基于关联——通过加工语境，语言使用者在心理词汇中激活了相应的语义和语音形式，从而对即将出现的词语形成预期(Kukona, 2020; 语音预测与语义预测对比见：Karimi et al., 2019)。值得注意的是，使用西方语言的语音预测研究存在一个无法避免的问题，即目标词(如 *cloud*)与其语音竞争项(如 *clown*)不仅在语音上有重合，在正字法信息上也存在交叠。Li 等人(2022)使用语音与正字法信息相对分离的汉语，通过视觉情境范式实验，也发现了类似的语音形式预测，验证了语音预测的普遍性。

3.2 句法加工的解歧过程

视觉情境范式对于句法加工时间进程研究的贡献主要在两个方面。首先，该范式可以用于分析歧义句的解歧过程，如花园路径句(garden-path sentences)。Tanenhaus 等人(1995)首次采用视觉情境范式探究了存在结构歧义的英文句子加工过程，以及视觉场景对句子解歧的影响。如 *Put the apple on the towel in the box* 在 *in the box* 出现前存在结构歧义：*on the towel* 既可以是动作 *put* 的方向，又可以是 *the apple* 的地点限定语。采用视觉情境范式眼动追踪的实验方法，Tanenhaus 等人发现在视觉场景中只有一个苹果的时候，听者会更倾向于把 *on the towel* 解读为动作的方向(眼动注视从苹果直接移向毛巾)；而当视觉场景中有两个苹果时，听者则更倾向于将其解读为 *the apple* 的地点限定语而非动作方向(在锁定毛巾上的苹果之后直接看向真正的目标地点——*the box* 箱子)。

其次，视觉情境范式为句法加工中不同层面信息何时被加工这一问题提供了新的证据。早期的双阶段理论(two-stage account)认为在句子理解过程中，句法结构分析要先于其他非结构性信息(包括词汇语义、世界知识、语篇等)的加工(initial syntactic analysis, Frazier, 1987)；基于约束的语言加工理论(constraint-based account)则认为句子加工涉及到多个层面信息的共同限制(Trueswell et al., 1994)，这些限制会在句子加工的早期就对句法结构分析产生影响。

视觉情境范式实验研究支持了后者的假说。如：Snedeker 和 Trueswell (2004)研究了具有歧义的介词短语结构(*Choose the cow with the stick* vs *Tickle the pig with the fan*)。*With the stick/fan* 既可以是宾语的限定成分，又可以是完成动作所借助的工具。他们发现，视觉场景中的信息(物体的个数)、动词的偏向(偏向限定语解读的动词 *choose*“选择” vs 偏向动作工具解读的动词 *tickle*“挠”)都会在句子加工的早期对歧义句的句法结构的分析产生影响，体现在物体个数、动词偏向不同的情况下，听者会看向不同的目标对象。此外，Chambers 等人(2002, 2004)的研究还发现，与视觉场景中物体形态、大小、特质相关的世界知识信息也会影响句法结构的分析，并且这些影响都发生在句子加工的最开始阶段，驳斥了句法结构分析为先的理论性假设。

3.3 语义的预测性加工

视觉情境范式对语义加工研究的一大贡献是，揭示了语义加工不仅是即时的，在很多情况下甚至是具有预测性的(Altmann & Kamide, 1999; Kamide et al., 2003; 理论综述见：Pickering & Gambi, 2018)。Altmann 和 Kamide (1999)最早使用视觉情境范式，研究了动词-论元整合的时间进程：与无关动词 move“移动”相比，听者在听到 *the boy will eat*…的动词 *eat* “吃”时，会更早地注视到视觉场景中的蛋糕这一物体上。这说明动词的语义信息(即 *eat*“吃”需要搭配可以吃的论元)会帮助听者预测论元的指代对象。Kamide 等人(2003)的后续研究总结了语义加工的主要特征：(1)动词与主语的组合共同促进了语义预测，例如主语 *the man*“男人”与动词 *ride*“骑”的组合会预测高可能性宾语 *motorbike*“摩托车”；(2)除了动词之外，附着于论元的格标记也会激活预测加工，如在动词后置的日语中，听者在动词还未出现之前也可以通过格标记提前预测即将出现的论元指代对象。

使用视觉情境范式对语义加工的研究不仅限于动词-论元结构。Chow 和 Chen (2020)使用该范式研究了汉语量词信息与语境中世界知识的整合加工，发现汉语使用者可以根据语境中的世界知识，在加工的早期对将要出现的名词形成预期，而这种预期会受到量词的影响，在加工后期进一步修正。此外，Grüter 等人(2020)对一语者和二语者量词加工的研究发现，一语者与二语者都对量词包含的语法搭配信息敏感，并且会利用该信息进行预测性加工。但是，二语者在加工中会更加依赖语义信息(如：量词“条”会搭配长条状物体)，表现为当视觉场景中出现不符合量词语法搭配、但符合长条状语义的干扰物时，二语者会更多地注视干扰物。

3.4 语篇层面加工

视觉情境范式可以用于探究语篇理解的两个重要议题——指代关系与连接关系。首先，

视觉情境范式下的眼动追踪可以有效检验代词与先行词之间指代关系的建立过程。一般认为，当听者听到与前文语篇有共同指代关系的代词、并注视某相关物体时，可以说明此物体被认为是潜在的目标指代物(Runner et al., 2003)。基于这一机制，研究者利用视觉情境范式探讨了诸多指代关系加工中的时间进程问题。例如，Arnold 等人(2000)最早发现性别线索和指代对象被提及的顺序都对指代消解有即时性影响：听者可以在加工早期利用不同性的语言标记形式(如：英语单数第三人称 *he* 或者 *she*)锁定指代的目标；同时，句中第一位提及的人物(如：SVO 语序句子中的主语)会更容易被解读为指代对象。在针对隐含因果对代词消解影响的研究中，Pyykkönen 和 Järvikivi (2010)发现，隐含因果效应在动词之后就已经立刻显现，听者听到动词后会更多地注视动词所偏向的指代对象，如：在 *John frightened Bill because*...中，动词 *frighten*"惊吓"更偏向第一个人物，所以当听者听到 *frightened* 时，会更多地注视 *John*；而在 *John feared Bill because*...中，动词 *feared* "害怕"则更偏向第二个人物，当动词出现时，听者更多注视 *Bill*。这一发现证明了指代加工是即时发生的，甚至具有预测性，而非延迟整合(另见：Cozijn et al., 2011)。

视觉情境范式也为连接关系在实时语言理解中的建立提供了丰富的实证证据。Wei 等人(2019)采用视觉情境范式探究了主观因果关系(论点-论据)和客观因果关系(原因-结果)的加工以及汉语连词在其中的作用。研究发现，相较于客观因果关系连词"因而"，当听者听到标记主观因果关系的连词"可见"时，相对于客观因果关系连词"因而"，他们会更多地注视视觉场景中的说话人。这表明主观与客观因果关系的加工可能在确认、追踪说话人的过程上有所不同，而且追踪说话人的过程是随着主观因果连词的输入而即时发生的，实验证据证明了语篇加工的即时性。Mak 等人(2017)通过在视觉场景中提供两个备选的指代对象，并追踪听者对两个指代对象的注视轨迹，探究俄语的两个连词在连接关系建立中的作用。研究发现，连词 *i*"而且"(用于标记延续关系，连词前后两个从句的主语一致)和连词 *a*"而且/但是"(用于标记转变关系，前后两个从句是不同的主语)可以帮助单语儿童和双语儿童提前预测第二个从句的主语是否转变，印证了在口语语篇理解中存在的预测性加工现象。

3.5 语用信息的提取与加工

语用隐含义(pragmatic implicature)何时被加工、这一过程是否先于语义分析是语用学领域关注的重要议题。字面义先行假设(literal-first hypothesis; Huang & Snedeker, 2009, 2011)认为对等级含义词字面语义(如 *some*"一些"的语义解读应为：一些-同时可以是全部)的加工先于该词的语用隐含义(一些-但并非全部)；Levinson (2000)认为语用隐含义是默认自动加工的；基于约束的加工理论则认为语用隐含义是否优先激活取决于是否具有充足的语境支持

(Degen & Tanenhaus, 2015, 2016)。

视觉情境范式是对比语义和语用信息加工时间线的重要实验手段。Huang 和 Snedeker (2011)的视觉情境范式眼动实验发现，听者在加工 *some*“一些”时会先注视与 *some* 语义解读(一些-同时可以是全部)相符的对象，而利用 *some*“一些”的语用隐含义(一些-但并非全部)来消除歧义、排除 *all*“全部”的指代对象这一过程要晚于 *some* 的语义加工(约晚 800 毫秒)。Degen 和 Tanenhaus (2016)的研究则发现，语用隐含义加工延迟的现象仅仅出现在当数字词也作为指令出现的情况下；而当数字词不存在时，*some* 的语用隐含义加工并不会晚于字面语义含义的加工。Gardner 等人(2021)改进了 Huang 和 Snedeker (2011)实验中的视觉物体个数使其更加符合 *some* 的概念，他们发现当有足够的语境支持时，语用隐含义的加工是迅速即时的，即听者可以运用 *some* 的语用隐含义快速锁定目标对象。此外，语言使用者对语用信息的加工还很大程度受到说话人可信度的影响——面对可信度高的说话人，受试者可以较早地利用等级形容词的语用含义锁定目标物体，而面对可信度低的说话人，则未出现早期的语用加工效应(Gardner et al., 2021)。

4. 视觉情境范式的主要贡献、局限性与研究展望

眼动视觉情境范式为研究语言理解提供了两项重要信息：一是视觉维度的注视指标；二是精确的时间测量。前者为心理语言学、认知心理学等领域的实验设计提供了丰富的可能性；而精确的时间测量，为语音、词汇、句法、语义、语篇、语用等各个层面的口语加工提供了准确的时间进程信息，极大地拓展了语言理解的相关理论。两者结合，可以有效反映在接收到口语信息输入时，听者在视觉场景中的注视位置如何随着时间变化，进而为语言理解中的一项重要议题——口语加工的时间进程提供了直接证据。视觉情境范式的实验研究通过分析高时间敏感性的眼动测量数据，发现语言各个层面的加工都呈现出即时性甚至预测性的特点，这与一些早期研究中语言延时整合的发现不同，说明语言加工时间进程的研究结果与所采用的方法密不可分。此外，视觉情境范式主要依赖听力任务，并不需要受试者具有完整的识字阅读能力，可以用来考察低龄儿童、二语学习者、特殊语言障碍人群的语言加工过程(研究示例见：Canseco-Gonzalez et al., 2010; McMurray et al., 2010; Weber & Cutler, 2004)。

视觉情境范式的主要局限性之一在于无法提供加工时长的数据，因此不能解答语言理解加工困难的相关问题(Salverda & Tanenhaus, 2018)。而且视觉情境范式实验只能在视觉空间中呈现数目有限的静态物体，这也与日常语言理解的复杂视觉环境有所区别。真实的语言理解环境可能包括更多的物体以及动态的动作、事件等，这也导致了该范式获得的结果在可推

广性上有一定局限(Huettig et al., 2011)。此外，在只呈现有限数目物体的实验环境下，听者可能会提前对语言输入形成一定的预期，并策略性地注视某些物体，因此眼动注视轨迹可能并不完全反映语言加工的过程 (Henderson & Ferreira, 2004)。对于这点质疑，Dahan 和 Tanenhaus (2004)根据其在词汇识别上的研究提出了不同意见，他们发现词频对词汇识别的影响效应并不会受到视觉空间中是否存在竞争项以及竞争项数目的影响，由此推断在视觉空间中提供有限数目的物体这一设置并不会影响视觉情境范式的有效性。

视觉情境范式的眼动研究仍有很大的发展空间。首先，尽管连接假设理论中所提出的关于视觉信息和语言信息的理解过程假设已经被大量实证结果所证实，任务目标对语言加工的重要作用仍然有待进一步探究。对比不同任务目标下，语言的加工过程如何随着时间发展，将是未来视觉情境范式眼动研究的方向之一。近年来，眼动研究也开始使用三维虚拟现实(VR)技术，这一技术创新可以高度还原自然的语言交流场景，同时保持对实验设置的精确控制。一些利用 VR 技术的视觉情境范式眼动实验，成功复现了语言加工中的一些经典结果，如预测性语言加工 (Eichert et al., 2018; Heyselaar et al., 2020)。这类技术改进不仅提高了视觉情境范式的生态效度，还可以用于检验在接近真实语言使用环境时，影响语言加工过程的诸多因素。理论和技术的创新都为更准确有效地收集解读眼动数据、探索语言加工提供了新的契机与更多的可能性。

参考文献

林桐，王娟．(2018). 基于视觉情境范式的口语词汇理解研究进展. *心理技术与应用*,*6*(09):570-576.

邱丽景，王穗苹，关心．(2009). 口语理解的视觉－情境范式研究. *华南师范大学学报*, *1*, 130-136.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *22*(1), 1–12. https://doi.org/10.1016/j.cub.2009.12.014

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The “blank screen paradigm.” *Cognition*, *93*(2), 79–87. https://doi.org/10.1016/j.cognition.2004.02.005

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the

domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/s0010-0277(99)00059-1

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518. https://doi.org/10.1016/j.jml.2006.12.004

Altmann, G. T. M., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, *33*(4). https://doi.org/10.1111/j.1551-6709.2009.01022.x.

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition*, *76*(1), B13–B26. https://doi.org/10.1016/s0010-0277(00)00073-1

Barr, D. J., Jackson, L., & Phillips, I. (2014). Using a voice to put a name to a face: The psycholinguistics of proper name comprehension. *Journal of Experimental Psychology: General*, *143*(1), 404–413. https://doi.org/10.1037/a0031813

Canseco-Gonzalez, E., Brehm, L., Brick, C. A., Brown-Schmidt, S., Fischer, K., & Wagner, K. (2010). Carpet or cárcel: The effect of age of acquisition and language mode on bilingual lexical access. *Language and Cognitive Processes*, *25*(5), 669–705. https://doi.org/10.1080/01690960903474912

Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language*, *47*(1), 30–49. https://doi.org/10.1006/jmla.2001.2832

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning Memory and Cognition*, *30*(3), 687–696. https://doi.org/10.1037/0278-7393.30.3.687

Chow, W. Y., & Chen, D. (2020). Predicting (in)correctly: Listeners rapidly use unexpected information to revise their predictions. *Language, Cognition and Neuroscience*, *35*(9), 1149–1161. https://doi.org/10.1080/23273798.2020.1733627

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken Language. *Cognitive Psychology*, *107*(1), 84–107. https://doi.org/10.1016/0010-0285(74)90005-x

Corps, R. E., Brooke, C., & Pickering, M. J. (2021). Prediction involves two stages: Evidence

from visual-world eye-tracking. *Journal of Memory and Language*, *122*, 104298. https://doi.org/10.1016/j.jml.2021.104298

Cozijn, R., Commandeur, E., Vonk, W., & Noordman, L. G. . (2011). The time course of the use of implicit causality information in the processing of pronouns: A visual world paradigm study. *Journal of Memory and Language*, *64*(4), 381–403. https://doi.org/10.1016/j.jml.2011.01.001

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning Memory and Cognition*, *30*(2), 498–513. https://doi.org/10.1037/0278-7393.30.2.498

Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicature: A constraint-based approach. *Cognitive Science*, *39*(4), 667–710. https://doi.org/10.1111/cogs.12171

Degen, J., & Tanenhaus, M. K. (2016). Availability of alternatives and the processing of scalar implicatures: A visual world eye-tracking study. *Cognitive Science*, *40*(1), 172–201. https://doi.org/10.1111/cogs.12227

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–1121. https://doi.org/10.1038/nn1504

Eichert, N., Peeters, D., & Hagoort, P. (2018). Language-driven anticipatory eye movements in virtual reality. *Behavior Research Methods*, *50*(3), 1102–1115. https://doi.org/10.3758/s13428-017-0929-z

Ferreira, F., Foucart, A., & Engelhardt, P. E. (2013). Language processing in the visual world: Effects of preview, visual complexity, and prediction. *Journal of Memory and Language*, *69*(3), 165–182. https://doi.org/10.1016/j.jml.2013.06.001

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 559–586). Lawrence Erlbaum Associates.

Garnham, A., Traxler, M., Oakhill, J., & Gernsbacher, M. A. (1996). The locus of implicit causality effects in comprehension. *Journal of Memory and Language*, *35*(4), 517–543. https://doi.org/doi.org/10.1006/jmla.1996.0028

Gardner, B., Dix, S., Lawrence, R., Morgan, C., Sullivan, A., & Kurumada, C. (2021). Online

pragmatic interpretations of scalar adjectives are affected by perceived speaker reliability. *PLoS ONE*, *16*(2), e0245130. https://doi.org/10.1371/journal.pone.0245130

Grüter, T., Lau, E., & Ling, W. (2020). How classifiers facilitate predictive processing in L1 and L2 Chinese: The role of semantic and grammatical cues. *Language, Cognition and Neuroscience*, *35*(2), 221–234. https://doi.org/10.1080/23273798.2019.1648840

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science*, *28*(1), 105–115. https://doi.org/10.1016/j.cogsci.2003.10.002

Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists. In J. M. Henderson & F. Ferreira (Eds.), *The Interface of Language, Vision, and Action: Eye Movements and the Visual World* (pp. 1–58). Psychology Press. https://doi.org/10.4324/9780203488430

Henry, N., Jackson, C. N., & Hopp, H. (2022). Cue coalitions and additivity in predictive processing: The interaction between case and prosody in L2 German. *Second Language Research*, *38*(3), 397–422. https://doi.org/10.1177/0267658320963151

Heyselaar, E., Peeters, D., & Hagoort, P. (2020). Do we predict upcoming speech content in naturalistic environments? *Language, Cognition and Neuroscience*, *36*(4), 440–461. https://doi.org/10.1080/23273798.2020.1859568

Huang, Y., & Snedeker, J. (2020). Evidence from the visual world paradigm raises questions about unaccusativity and growth curve analyses. *Cognition*, *200*, 104251. https://doi.org/10.1016/j.cognition.2020.104251

Huang, Y. T., & Snedeker, J. (2009). Semantic meaning and pragmatic interpretation in 5-year-olds: Evidence from real-time spoken language comprehension. *Developmental Psychology*, *45*(6), 1723–1739. https://doi.org/10.1037/a0016704

Huang, Y. T., & Snedeker, J. (2011). Logic and conversation revisited: Evidence for a division between semantic and pragmatic content in real-time language comprehension. *Language and Cognitive Processes*, *26*(8), 1161–1172. https://doi.org/10.1080/01690965.2010.508641

Huettig, F., & Guerra, E. (2019). Effects of speech rate, preview time of visual context, and participant instructions reveal strong limits on prediction in language processing. *Brain Research*, *1706*, 196–208. https://doi.org/10.1016/j.brainres.2018.11.013

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape

information in language-mediated visual search. *Journal of Memory and Language*, *57*(4), 460–482. https://doi.org/10.1016/j.jml.2007.02.001

Huettig, F., Olivers, C. N. L., & Hartsuiker, R. J. (2011a). Looking, language, and memory: Bridging research from the visual world and visual search paradigms. *Acta Psychologica*, *137*(2), 138–150. https://doi.org/10.1016/j.actpsy.2010.07.013

Huettig, F., Rommers, J., & Meyer, A. S. (2011b). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. https://doi.org/10.1016/j.actpsy.2010.11.003

Ito, A., & Knoeferle, P. (2022). Analysing data from the psycholinguistic visual-world paradigm: Comparison of different analysis methods. *Behavior Research Methods*. https://doi.org/10.3758/s13428-022-01969-3

Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language*, *98*, 1–11. https://doi.org/10.1016/j.jml.2017.09.002

Kaan, E., & Grüter, T. (2021). Prediction in second language processing and learning: Advances and directions. In E. Kaan & T. Grüter (Eds.), *Prediction in second language processing and learning* (pp. 1–24). John Benjamins.

Kaiser, E. (2016). Discourse-level Processing. In P. Knoeferle, P. Pyykkönen-Klauck, & M. W. Crocker (Eds.), *Visually situated language comprehension* (pp. 151–184). John Benjamins Publishing.

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive procesing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, *32*(1), 37–55. https://doi.org/10.1023/a:1021933015362

Karimi, H., Brothers, T., & Ferreira, F. (2019). Phonological versus semantic prediction in focus and repair constructions: No evidence for differential predictions. *Cognitive Psychology*, *112*, 25–47. https://doi.org/10.1016/j.cogpsych.2019.04.001

Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science*, *30*(3), 481–529. https://doi.org/10.1207/s15516709cog0000_65

Knoeferle, P., & Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: Evidence from eye movements. *Journal of Memory and Language*, *57*(4), 519–543. https://doi.org/10.1016/j.jml.2007.01.003

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, *95*(1), 95–127. https://doi.org/10.1016/j.cognition.2004.03.002

Koornneef, A. W., & Van Berkum, J. J. A. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language, 54*, 445–465. https://doi.org/10.1016/j.jml.2005.12.003

Koring, L., Mak, P., & Reuland, E. (2012). The time course of argument reactivation revealed: Using the visual world paradigm. *Cognition*, *123*(3), 361–379. https://doi.org/10.1016/j.cognition.2012.02.011

Kukona, A. (2020). Lexical constraints on the prediction of form: Insights from the visual world paradigm. *Journal of Experimental Psychology: Learning Memory and Cognition*, *46*(11), 2153–2162. https://doi.org/10.1037/xlm0000935

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language*, *Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT Press.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Li, X., Li, X., & Qu, Q. (2022). Predicting phonology in language comprehension: Evidence from the visual world eye-tracking task in Mandarin Chinese. *Journal of Experimental Psychology: Human Perception and Performance*, *48*(5), 531–547. https://doi.org/10.1037/xhp0000999

Magnuson, J. S. (2019). Fixations in the visual world paradigm: Where, when, why? *Journal of Cultural Cognitive Science*, *3*(2), 113–139. https://doi.org/10.1007/s41809-019-00035-3

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken

word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General*, *132*(2), 202–227. https://doi.org/10.1037/0096-3445.132.2.202

Mak, W. M., Tribushinina, E., Lomako, J., Gagarina, N., Abrosova, E., & Sanders, T. (2017). Connective processing by bilingual children and monolinguals with specific language impairment: Distinct profiles. *Journal of Child Language*, *44*(2), 329–345. https://doi.org/10.1017/s0305000915000860

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, *53*(4), 372–380. https://doi.org/10.3758/bf03206780

McMurray, B., Samelson, V. M., Lee, S. H., & Tomblin, J. B. (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology*, *60*(1), 1–39. https://doi.org/10.1016/j.cogpsych.2009.06.003

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, *38*(3), 283–312. https://doi.org/10.1006/jmla.1997.2543

Millis, K. K., & Just, M. A. (1994). The influence of connectives on sentence comprehension. *Journal of Memory and Language*, *33*(1), 128–147. https://doi.org/10.1006/jmla.1994.1007

Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R*. CRC Press.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*(4), 475–494. https://doi.org/10.1016/j.jml.2007.11.006

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., Von Grebmer Zu Wolfsthurn, S., Bartolozzi, F., Kogan, V., Ito, A., Mézière, D., Barr, D. J., Rousselet, G. A., Ferguson, H. J., Busch-Moreno, S., Fu, X., Tuomainen, J., Kulakova, E., Husband, E. M., … Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *ELife*, *7*, 1–24. https://doi.org/10.7554/eLife.33468

Porretta, V., Kyröläinen, A. J., Rij, J. Van, & Järvikivi, J. (2018). Visual world paradigm data: From preprocessing to nonlinear time-course analysis. In I. Czarnowski, R. Howlett, & L. Jain (Eds.), *Intelligent Decision Technologies 2017. Smart Innovation, Systems and*

Technologies (Vol. 73, pp. 268–277). Springer.

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002–1044. https://doi.org/10.1037/bul0000158

Pyykkönen-Klauck, P., & Crocker, M. W. (2016). Attention and eye movement metrics in visual world eye tracking. In P. Knoeferle, P. Pyykkönen-Klauck, & M. W. Crocker (Eds.), *Visually Situated Language Comprehension* (pp. 67–82). John Benjamins Publishing.

Pyykkönen, P., & Järvikivi, J. (2010). Activation and persistence of implicit causality information in spoken language comprehension. *Experimental Psychology*, *57*(1), 5–16. https://doi.org/10.1027/1618-3169/a000002

Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin*, *85*(3), 618–660. https://doi.org/10.1037/0033-2909.85.3.618

Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2003). Assignment of reference to reflexives and pronouns in picture noun phrases: Evidence from eye movements. *Cognition*, *89*, B1–B13. https://doi.org/10.1016/S0010-0277(03)00065-9

Salverda, A. P., Brown, M., & Tanenhaus, M. K. (2011). A goal-based perspective on eye movements in visual world studies. *Acta Psychologica*, *137*(2), 172–180. https://doi.org/10.1016/j.actpsy.2010.09.010

Salverda, A. P., & Tanenhaus, M. K. (2018). The visual world paradigm. In A. M. B. de Groot & P. Hagoort (Eds.), *Research Methods in Psycholinguistics and the Neurobiology of Language: A Practical Guide* (pp. 89–110). Wiley-Blackwell.

Saryazdi, R., & Chambers, C. G. (2021). Real-time communicative perspective taking in younger and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(3), 439–454.

Saslow, M. G. (1967). Latency of saccadic eye movement. *Journal of the Optical Society of America*, *57*(8), 1030–1033. https://doi.org/10.2466/pms.2003.96.1.173

Seedorff, M., Oleson, J., & McMurray, B. (2018). Detecting when timeseries differ: Using the bootstrapped differences of timeseries (BDOTS) to analyze visual world paradigm data (and more). *Journal of Memory and Language*, *102*, 55–67. https://doi.org/10.1016/j.jml.2018.05.004

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role

of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*(3), 238–299. https://doi.org/10.1016/j.cogpsych.2004.03.001

Stewart, A. J., Pickering, M. J., & Sanford, A. J. (2000). The time course of the influence of implicit causality information: Focusing versus integration accounts. *Journal of Memory and Language*, *42*(3), 423–443. https://doi.org/10.1006/jmla.1999.2691

Stone, K., Lago, S., & Schad, D. J. (2021). Divergence point analyses of visual world data: Applications to bilingual research. *Bilingualism: Language and Cognition*, *24*(5), 833–841. https://doi.org/10.1017/s1366728920000607

Tanenhaus, M. K., Magnuson, J. S., & Dahan, D. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research*, *29*(6), 557–580. https://doi.org/10.1023/a:1026464108329

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634. https://doi.org/10.1126/science.7777863

Traxler, M. J., Bybee, M. D., & Pickering, M. J. (1997). Influence of connectives on language comprehension: Eye tracking evidence for incremental interpretation. *The Quarterly Journal of Experimental Psychology*, *50A*(3), 481–497. https://doi.org/10.1080/027249897391982

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*(3), 285–318. https://doi.org/10.1006/jmla.1994.1014

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*(1), 1–25. https://doi.org/10.1016/S0749-596x(03)00105-0

Wei, Y., Mak, W. M., Evers-Vermeul, J., & Sanders, T. J. M. (2019). Causal connectives as indicators of source information: Evidence from the visual world paradigm. *Acta Psychologica*, *198*, 102866. https://doi.org/10.1016/j.actpsy.2019.102866

Visual world paradigm reveals the time course of spoken language processing

Yipu WEI　School of Chinese as a Second Language, Peking University

Abstract: The visual world paradigm (VWP) assesses real-time language processing by tracking and measuring eye movements in visual contexts. Linking hypotheses, such as the coordinated interplay account and the goal-based linking hypothesis, establish the link between eye movements and the cognitive processes of language comprehension. Time sensitivity is characteristic of the data generated by this paradigm. Analytical methods include the analysis of fixation proportions within time windows, divergence point analysis and growth-curve analysis, etc. Studies using the VWP provide important evidence for speech and lexical recognition, syntactic parsing, semantic integration, and the processing of discourse and pragmatic information.

Keywords: visual world paradigm; eye-tracking; spoken language processing